DAWAN

Plan de la formation

www.dawan.fr

Formation Talend Open Studio Data Integration : Initiation + Approfondissement

Durée :	5 jours (35 heures)
Tarifs inter-entreprise :	2 995,00 CHF HT (standard) 2 396,00 CHF HT (remisé)
Public :	Administrateurs systèmes/BDD, Chefs de projets, Développeurs
Pré-requis :	Notions de Java et fondamentaux SQL
Objectifs :	Comprendre le processus d'intégration de données - Maîtriser l'outil Talend Open Studio et la bibliothèque de composants
Modalités pédagogiques, techniques et d'encadrement :	• Formation synchrone en présentiel et distanciel. • Méthodologie basée sur l'Active Learning : 75 % de pratique minimum. • Un PC par participant en présentiel, possibilité de mettre à disposition en bureau à distance un PC et l'environnement adéquat. • Un formateur expert.
Modalités d'évaluation :	• Définition des besoins et attentes des apprenants en amont de la formation. • Auto-positionnement à l'entrée et la sortie de la formation. • Suivi continu par les formateurs durant les ateliers pratiques. • Évaluation à chaud de l'adéquation au besoin professionnel des apprenants le dernier jour de formation.
Sanction :	Attestation de fin de formation mentionnant le résultat des acquis
Référence :	BUS305-F
Note de satisfaction des participants:	4,73 / 5

Contacts :	commercial@dawan.fr - 09 72 37 73 73
Modalités d'accès :	Possibilité de faire un devis en ligne (www.dawan.fr, moncompteformation.gouv.fr, maformation.fr, etc.) ou en appelant au standard.
Délais d'accès :	Variable selon le type de financement.
Accessibilité :	Si vous êtes en situation de handicap, nous sommes en mesure de vous accueillir, n'hésitez pas à nous contacter à referenthandicap@dawan.fr, nous étudierons ensemble vos besoins

Introduction

Problématique d'intégration de données
Intégration Opérationnelle et Décisionnelle
Outils ETL (Extract Transform Load) disponibles
Talend Open Studio : présentation, licences, lien avec Eclipse
Support disponible : documentation, communauté, ...

Atelier : Installation et présentation de l'environnement Talend Open Studio

Découvrir l'interface

Les perspectives et les vues Talend
Utilisation du Business Modeler
Les jobs : grille, palette de composants
Les flux principaux (main)
La notion de schéma : types des colonnes
Premiers composants : entrée, sortie, génération de flux

Atelier pratique : création et exécution d'un job simple, génération de contenu dans un fichier

Les principaux composants de Talend

Composants d'entrée (tFileInput, tMysqlInput, ...)
Génération de contenu (tRowGenerator, tFixedFlowInput, ...)
Sorties et affichage (tLogRow, tFileOutput, tMysqlOutput, ...)
Logique et chaînage des composants
Suivi pas à pas des traces de l'exécution du job
Gestion des rejets
Transformations simples (filtre, tri, remplacement, jointure)

Atelier pratique : génération de contenu dans une base de données, extraction et traitement des données pour affichage en console

Garantir la pérennité des jobs

Versions et statuts des jobs
Les métadonnées : propriétés et schémas.
La découverte de schéma pour les fichiers CSV
Récupération des schémas de bases de données
Mises à jour des jobs depuis les métadonnées

Atelier pratique : insérer les sources et cibles de données dans les métadonnées du référentiel

Variables et contextes

Utiliser des variables dans un job (tSetGlobalVar)
Mise en place d'un contexte local au job
Groupes de contexte dans le référentiel
Les métadonnées comme contextes

Atelier pratique : utiliser une variable dans un job, contextualiser les paramètres d'une métadonnées, créer un contexte de test, et de production

Le mappage de données : tMap

Le composant ETL par excellence
Gérer les entrées : flux principal, secondaires, nommage des flux
Gérer les sorties : l'éditeur intégré de la tMap
Filtrer les données : en entrée, sortie
Sorties conjointes, capture des rejets de sortie, de jointure interne
L'éditeur d'expression
Les performances

Atelier pratique : extraction multi-sources, multi-destinations avec une tMap, sur un exemple complet

Orchestration des jobs, gestion d'erreurs

Exécutions parallèles de sous-jobs
Déclencheurs Talend (triggers)
Enchaînement de sous-jobs et de composants (OnSubjobOK, OnComponentOK)

Gestion des erreurs de composants, sous-job (OnComponentError, OnSubjobError)
Composant de départ, et de fin (tPreJob, tPostJob)
Attraper et lever des erreurs : tWarn, tDie, tLogCatcher

Atelier pratique : assembler plusieurs jobs en un seul, factoriser la gestion d'erreurs pour générer un fichier de compte-rendu

Les flux de programmation : liens iterate

Différencier les flux principaux des flux iterate
Boucles et programmation via tLoop, tForeach
Passer d'un flux principal à un flux iterate, et l'inverse (tFlowToIterate, tIterateToFlow)
Synchronisation de jobs avec les tWaitForFile, tWaitForSocket, tWaitForSQL

Atelier pratique : paralléliser des sous-jobs, synchroniser les jobs avec un événement externe

Déploiement et industrialisation

Rappel sur les contextes
Export des jobs pour exécution sur un serveur
Passage de contexte entre jobs, à l'exécution
Utilisation d'un fichier pour charger le contexte (tContextLoad)

Atelier : Déploiement de tâches et chargement dynamique du contexte

Utilisation avancée

Utilisation de composants évolués : Java ou expressions régulières
Création de logs spécifiques
Gestion des flux XML
Réutilisation du code : les routines

Atelier : Utilisation des composants tJavaRow, tJavaFlex - création de logs avec tCatch, tDie ... - utilisation de l'éditeur Java embarqué

Les composants Talend

Concepts essentiels pour étendre Talend Open Studio
Création de la maquette du composant avec tJavaFlex / interface graphique
Découvrir la forge Talend et ses composants communautaires
Intégration dans Talend Open Studio des plugins créés ou téléchargés

Atelier : Création d'un composant simple via tJavaFlex, intégration de composants de la forge Talend